

Perfect Sampling of Markov Chains with Piecewise Homogeneous Events

Ana BUŠIĆ* Bruno GAUJAL† Furcy PIN*

Abstract

Perfect sampling is a technique that uses coupling arguments to provide a sample from the stationary distribution of a Markov chain in a finite time without ever computing the distribution. This technique is very efficient if all the events in the system have monotonicity property. However, in the general (non-monotone) case, this technique needs to consider the whole state space, which limits its application only to chains with a state space of small cardinality. We propose here a new approach for the general case that only needs to consider two trajectories. Instead of the original chain, we use two bounding processes (envelopes) and we show that, whenever they couple, one obtains a sample under the stationary distribution of the original chain. We show that this new approach is particularly effective when the state space can be partitioned into pieces where envelopes can be easily computed. We further show that most Markovian queueing networks have this property and we propose efficient algorithms for some of them.

Keywords: Markov chains, perfect sampling, queueing systems.

AMS classification (2010): Primary: 60J10, 65C05. Secondary: 68M20.

1 Introduction

Simulation approaches can be efficient ways to estimate the stationary behavior of Markov chains by providing independent samples distributed according to their stationary distributions, even when it is impossible to compute this distribution numerically.

Propp and Wilson used a backward coupling scheme [14] to derive a simulation algorithm - called PSA (Perfect Sampling Algorithm) in the following - providing perfect samples (*i.e.* whose distribution is stationary) of the state of a *discrete time Markov chain* over a finite state space \mathcal{S} . The

*INRIA/ENS, 23, avenue d'Italie, CS 81321, 75214 Paris Cedex 13, France. E-mail: Ana.Busic@inria.fr, pin@clipper.ens.fr.

†INRIA and LIG, 51, avenue Jean Kuntzmann Montbonnont, France. E-mail: bruno.gaujal@inria.fr.

main idea is to simulate trajectories starting from all states $x \in \mathcal{S}$ at some time in the past until time $t = 0$. If the state at time 0 is the same for all trajectories, then the chain has *coalesced* and the final state has the stationary distribution of the Markov chain. Otherwise, the simulations are started further in the past. Perfect sampling procedures have been developed in various contexts (for more information, see the annotated bibliography [16]).

Although the Perfect Sampling Algorithm provides perfect samples for all irreducible and aperiodic finite Markov chains, its complexity can be high. It suffers from two drawbacks that jeopardize its applicability for very large chains.

- The first one is the fact that the coupling time can be very large. Some recent work focused on the estimation of the coupling time for certain classes of Markov chains. For example, it was shown in [4] that Markov chains, modeling a class of networks of queues with finite capacities, have a quadratic coupling time with respect to capacities.
- The second factor in the complexity of PSA is the fact that one needs to run one simulation per state in \mathcal{S} , which is prohibitive for most applications. Various techniques have been developed to reduce the number of trajectories that need to be considered in the coupling from the past procedure. A first crucial observation already appears in [14]: for a monotone Markov chain, one has to consider only extremal initial conditions. For anti-monotone systems, an analogous technique that also considers only extremal initial conditions has been developed by Kendall [11] and Häggström and Nelander [8].

To cope with non-monotonicity, Kendall and Møller introduced in [12] the general idea of two bounding processes sandwiching all the normal trajectories, reducing the number of processes having to be computed to two. They applied the sandwiching technique to perfect sampling of point processes, together with the idea of a dominating chain that allowed to handle the infinite state space. The idea of sandwiching processes was also applied in communication networks [1, 7]. Huber [9] also introduced a similar idea of bounding chains for determining when coupling has occurred. In his model, the bounding chains are evolving in some bigger state space than the target Markov chain. However, both Kendall-Møller's and Huber's constructions are model-dependent and do not have an algorithmic counterpart so that they are not straightforward to apply in general.

In this paper we introduce a new perfect sampling algorithm (EPSA: Envelope Perfect Sampling Algorithm) that uses the lattice structure of the state space to design an effective and automatic construction of a bounding interval chain that can be expressed using only two trajectories. This is done by computing from the past, a lower and an upper bound of all trajectories

of the Markov chain. Whenever the two bounds meet, the coupling state is distributed according to the stationary distribution of the chain. The idea of bounding interval chain is close to sandwiching processes in Kendall and Møller [12]. However, their approach was not constructive and the sandwiching processes were computed ad-hoc for each application.

Our aim is to propose, under a constructive approach, several algorithms to compute bounding interval chains that are easily computable and as tight as possible. This is done in two cases. The first (easy) case is when events are *almost state space homogeneous* (defined in Section 4). In that case, we provide an algorithm that computes the update rule of the bounding interval chain. The second case is more difficult since events are assumed to be piecewise defined (Section 5) over arbitrary polytopes (Section 6). In that case we show how to obtain an interval containing all trajectories by solving a set of linear programs. Since the worst case complexity of linear programs may be too high, we show how to avoid solving these, at the expense of obtaining bounding intervals that may be less tight.

In the last part of the paper (Section 7), we consider applications to queueing networks. We exhibit several events (such as negative customers, fork and join, batch routing) that satisfy almost state space homogeneity. We also provide numerical evidence of the very good performance of EPSA in these cases, by computing the coupling time experimentally.

We also show examples (such as generalized join the shortest queue) that are piecewise homogeneous, where the smallest interval containing all trajectories can also be computed in linear time in the number of queues. Our numerical experiments show that EPSA behaves just as well as in the easy case of almost space homogeneous events.

2 Definitions and Basic Properties

If not stated otherwise, all real (or integer) vectors are row vectors and v^t denotes the transposed vector. For $x, y \in \mathbb{R}^d$, we denote $x \vee y \stackrel{\text{def}}{=} (\max\{x_1, y_1\}, \dots, \max\{x_d, y_d\})$, and $x \wedge y \stackrel{\text{def}}{=} (\min\{x_1, y_1\}, \dots, \min\{x_d, y_d\})$. To avoid multiple brackets, for $x, y, z \in \mathbb{R}^d$, we will write:

$$x \vee y \wedge z \stackrel{\text{def}}{=} (x \vee y) \wedge z.$$

Note that $(x \vee y) \wedge z \neq x \vee (y \wedge z)$; the relations involving operators \vee and \wedge will be evaluated from left to right. For any $x \in \mathbb{R}^n$, $\|x\|_1 = \sum_{i=1}^n |x_i|$, and for any set U , $|U|$ denotes its cardinality.

Throughout the paper, we consider discrete time homogeneous Markov chains (DTMC) with a finite state space \mathcal{S} . Since we only consider finite state spaces, the continuous time models can be easily handled using the same techniques, after uniformization.

2.1 Markov Automaton

The evolution of a finite DTMC can always be obtained using a finite number of discrete events (or actions). To formalize this, we introduce the notion of Markov automaton. Markov automata are similar to Generalized Semi Markov Processes [3] with a focus on state changes rather than on time.

Definition 2.1. A Markov automaton is a tuple $\mathcal{A} = (\mathcal{S}, A, D, \cdot)$ where \mathcal{S} is a finite state space, A is an alphabet called the set of events, D is a probability distribution on A , and \cdot is a right action by the letters of A on the states in \mathcal{S} and is called the transition function :

$$\cdot : \mathcal{S} \times A \rightarrow \mathcal{S}, (x, a) \mapsto x \cdot a.$$

Equivalently, this action can be most naturally extended to a right action by words $u_{1 \rightarrow n} \stackrel{\text{def}}{=} u_1 \dots u_n$ in A^n on \mathcal{S} for any $n \in \mathbb{N}$:

$$\cdot : \mathcal{S} \times A^n \rightarrow \mathcal{S}, (x, u_{1 \rightarrow n}) \mapsto x \cdot u_{1 \rightarrow n} \stackrel{\text{def}}{=} x \cdot u_1 \cdot u_2 \cdot \dots \cdot u_n.$$

A Markov automaton $\mathcal{A} = (\mathcal{S}, A, D, \cdot)$ naturally induces a Markov chain with the following construction:

Let (u_1, \dots, u_n, \dots) be an infinite sequence of random letter of A *i.i.d.* distributed according to D . We consider this sequence as an infinite word $u \stackrel{\text{def}}{=} u_1 \dots u_n \dots$ on A . Then for any $x_0 \in \mathcal{S}$, the random process $(X_n \stackrel{\text{def}}{=} x_0 \cdot u_{1 \rightarrow n})_{n \in \mathbb{N}}$ is a Markov chain issued from x_0 with probability transition matrix P given by

$$\text{for all } x, y \text{ in } \mathcal{S}, \quad P(x, y) = \sum_{\substack{a \in A \\ x \cdot a = y}} \mathbb{P}_D(a). \quad (2.1)$$

We will say that the Markov chain (X_n) is *generated* by \mathcal{A} and u . Moreover, with this construction, we can build a family of Markov chains $\{(X_n(x) = x \cdot u_{1 \rightarrow n})_{n \geq 0} \mid x \in \mathcal{S}\}$ starting from each state $x \in \mathcal{S}$. We will refer to this family of Markov chain as the *grand coupling* generated by \mathcal{A} and u . We will say that the grand coupling has *coalesced* at time n if all the Markov chains of the family has reached the same state, *i.e.* when $\mathcal{S} \cdot u_{1 \rightarrow n}$ is reduced to a singleton.

Conversely, for any probability transition matrix P on a finite space state \mathcal{S} , it is easy to see that there exists a Markov automaton $\mathcal{A} = (\mathcal{S}, A, D, \cdot)$ such that (2.1) holds, *i.e.* such that \mathcal{A} generates a Markov chain on \mathcal{S} with same transition matrix P , but that automaton is in general not unique. However as we will see in our examples, such representation of Markovian systems with Markov automata naturally arises for many systems such as queueing networks.

2.2 Perfect Sampling

Let $\{X_n\}_{n \in \mathbb{N}}$ be an irreducible and aperiodic DTMC with a finite state space \mathcal{S} and a transition matrix P . Let π denote the steady state distribution of the chain: $\pi = \pi P$.

The Perfect Sampling Algorithm (PSA) was first used by Propp and Wilson. It allows one to draw in finite time steady state distributed variable using coupling from the past.

Consider a Markov automaton $\mathcal{A} = (\mathcal{S}, A, D, \cdot)$ and $(u_0, u_{-1}, \dots, u_{-n}, \dots)$ a sequence of events *i.i.d.* with distribution D on A .

Theorem 2.2 (Propp and Wilson [14]). *There exists $\ell \in \mathbb{N}$ such that*

$$\lim_{n \rightarrow \infty} |\mathcal{S} \cdot u_{-n+1 \rightarrow 0}| = \ell \text{ almost surely.}$$

The grand coupling generated by \mathcal{A} and u couples if $\ell = 1$. In that case, let $\tau \stackrel{\text{def}}{=} \inf \left\{ n : |\mathcal{S} \cdot u_{-n+1 \rightarrow 0}| = 1 \right\}$ be the coupling time of the chain. Then $\mathbb{E}(\tau) < \infty$ and $\mathcal{S} \cdot u_{-\tau+1 \rightarrow 0}$ is steady state distributed.

Algorithm 1: Perfect Sampling Algorithm (PSA)

Data: I.i.d. events $\{u_{-n}\}_{n \in \mathbb{N}} \in \mathcal{A}^{\mathbb{N}}$

Result: A state $x^* \in \mathcal{S}$ generated according to the stationary distribution of the Markov chain

begin

$n := 0$;

foreach state $x \in \mathcal{S}$ **do** $S[x] := x$;

 // $S[x]$ is the state of the trajectory issued from x at time $-n$.

repeat

foreach state $x \in \mathcal{S}$ **do** $R[x] := S[x \cdot u_{-n}]$; $S := R$;

$n := n + 1$;

until $|\{S[x], x \in \mathcal{S}\}| = 1$

 (x^* denotes this unique value);

return x^*

The algorithmic counterpart (PSA) of Theorem 2.2 is given in Algorithm 1. The main drawback of PSA algorithm is the fact that one needs to simulate one Markov chain starting from each state in \mathcal{S} , that could be too large for a practical use of the algorithm.

Several approaches have been used to overcome this problem. The main one for a partially ordered state space (\mathcal{S}, \preceq) and monotone events is already present in [14].

Definition 2.3. An event $a \in \mathcal{A}$ is said to be *monotone* if $x \preceq y \implies x \cdot a \preceq y \cdot a$. An event $a \in \mathcal{A}$ is said to be *anti-monotone* if $x \preceq y \implies x \cdot a \succeq y \cdot a$.

If all events are monotone, then it is possible to generate a steady state by starting Algorithm 1 with maximal and minimal states only [14]. This technique has been successfully used in [15] to construct PSA for network of queues. Similarly, an efficient perfect sampling algorithm for the chains with anti-monotone events was given in [8].

When events are neither monotone nor anti-monotone, one can still use monotone bounds, as in [5], but this technique provides a sample from a bounding chain and not the initial one. We give in this paper an algorithm for perfect sampling of general (non-monotone) chains.

3 Envelopes

The approach proposed here generalizes what has been done in [6] and [2] to sample non-monotone Markov chains. It uses a bounding chain method similar to the one employed by Kendall and Møller in [12]. The sandwiching chains in [12] are in general difficult to compute, except for some models for which effective ad-hoc rules can be derived. Our approach is more algorithmic. We introduce a general framework that provides efficient algorithms for a wide variety of events, including many queueing events.

3.1 Bounding Interval Chains

We assume from now on that the state space \mathcal{S} is equipped with a lattice order relation \preceq .

For $m, M \in \mathcal{S}$, we will denote by $[m, M] \stackrel{\text{def}}{=} \{x \in \mathcal{S} : m \preceq x \preceq M\}$ the *lattice interval* (or simply *interval*) between the endpoints m and M (note that $[m, M] \neq \emptyset$ if and only if $m \preceq M$). We introduce \mathcal{I} as the set of all nonempty lattice intervals: $\mathcal{I} = \{[m, M] : m, M \in \mathcal{S}, m \preceq M\}$.

We define bounding interval chains as follows.

Definition 3.1. Given a grand coupling $\{(X_n(x))_{n \geq 0} \mid x \in \mathcal{S}\}$, we call bounding interval chain of that grand coupling *any Markov chain of nonempty intervals* $([m_n, M_n])_{n \geq 0}$ such that for all x in \mathcal{S} and all $n \geq 0$, $X_n(x) \in [m_n, M_n]$.

In particular we notice that when $m_n = M_n$, the grand coupling inside has necessarily coalesced. Thus, by constructing efficiently such bounding intervals, we can detect easily if a grand coupling has coalesced and apply it to PSA.

We now construct the tightest bounding interval chain computable step by step for a given grand coupling, that we will call envelope chain.

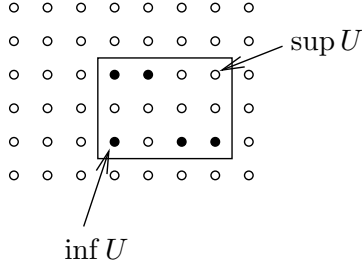


Figure 1: Intervals in dimension two for the natural product order. The set U is the set of all the black points. The interval $\llbracket U \rrbracket$ is displayed as a rectangle.

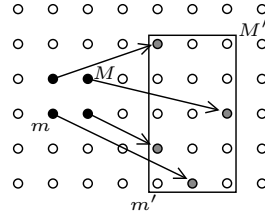


Figure 2: Envelopes of the action of a on an interval. The interval $[m, M]$ is displayed as black points. The action of a is displayed as arrows. The envelope interval $[m', M'] = [m, M] \sqcup a$ is displayed as a rectangle.

For any set $U \subset \mathcal{S}$, we denote by $\llbracket U \rrbracket$ and we call *envelope of U* the smallest interval of \mathcal{S} (for inclusion) containing U : $\llbracket U \rrbracket \stackrel{\text{def}}{=} [\inf(U), \sup(U)]$. This is illustrated by Figure 1. We introduce a new transition operator $\sqcup : \mathcal{I} \times \mathcal{A} \rightarrow \mathcal{I}$ called *envelope transition* that transforms intervals into intervals and is defined by : for all $[m, M] \in \mathcal{I}$ and $a \in \mathcal{A}$:

$$[m, M] \sqcup a \stackrel{\text{def}}{=} \llbracket [m, M] \cdot a \rrbracket = \left[\inf_{m \preceq x \preceq M} \{x \cdot a\}, \sup_{m \preceq x \preceq M} \{x \cdot a\} \right]. \quad (3.1)$$

The construction of \sqcup is illustrated on a simple example in Figure 2.

As with ‘ \cdot ’, we extend the operator to a finite word of events $u_{1 \rightarrow n} = u_1 \dots u_n \in \mathcal{A}^n$:

$$[m, M] \sqcup u_{1 \rightarrow n} \stackrel{\text{def}}{=} [m, M] \sqcup u_1 \sqcup u_2 \sqcup \dots \sqcup u_n.$$

Let us call $\perp \stackrel{\text{def}}{=} \inf \mathcal{S}$ (resp. $\top \stackrel{\text{def}}{=} \sup \mathcal{S}$) the bottom (resp. top) element of \mathcal{S} . The process

$$[m_n, M_n] \stackrel{\text{def}}{=} [\perp, \top] \sqcup u_{1 \rightarrow n}$$

is a Markov chain over the state space $\mathcal{S} \times \mathcal{S}$, called the *envelope chain*, and is a bounding interval chain of the grand coupling $\{(x \cdot u_{1 \rightarrow n}) \mid x \in \mathcal{S}\}$.

It is also the tightest bounding chain computable step by step, in the sense that if, at time n , we only know that $S \cdot u_{1 \rightarrow n}$ is in $[m_n, M_n]$, then the tightest interval containing $S \cdot u_{1 \rightarrow n+1}$ we can compute at time $n+1$ is $[m_n, M_n] \sqcup u_{n+1}$. However, it is not tight in the stronger sense of $[m_n, M_n]$ being equal to $\llbracket S \cdot u_{1 \rightarrow n} \rrbracket$, that does not hold in general.

Notice that the lower envelope $(m_n)_{n \in \mathbb{N}}$ alone is not a Markov chain, neither is the upper one $(M_n)_{n \in \mathbb{N}}$, since they depend on each other.

We can then use the envelope process to detect coupling in PSA.

Theorem 3.2. Assume that $[m_n, M_n]$ hits the set of single point intervals $\mathcal{X} = \{[x, x] : x \in \mathcal{S}\}$ in finite time:

$$\tau_e \stackrel{\text{def}}{=} \min \left\{ n : [\perp, \top] \boxdot u_{-n+1 \rightarrow 0} \in \mathcal{X} \right\},$$

then τ_e is a backward coupling time of the envelope chain. The state defined by $[\perp, \top] \boxdot u_{-\tau_e+1 \rightarrow 0}$ has the steady state distribution of DTMC $\{X_n\}_{n \in \mathbb{N}}$.

Proof. The proof simply uses the fact that the set $\mathcal{S} \cdot u_{-n+1 \rightarrow 0}$ is included in $[\perp, \top] \boxdot u_{-n+1 \rightarrow 0}$. Consequently, if the latter is reduced to one point, so is the former and the grand coupling has coalesced. \square

Algorithm 1 can be adapted to the envelope simulation: start the simulation with only two states, \perp and \top , and iterate using \boxdot and an *i.i.d.* sequence of events. This new algorithm, called EPSA (Envelope Perfect Sampling Algorithm), is given in Algorithm 2. The reason why we double n at each loop of the algorithm is that we need to compute $S \boxdot u_{-n+1 \rightarrow 0}$ in each loop, which corresponds to n iterations of \boxdot . While increasing n by 1 at each loop would lead to a quadratic cost in n , doubling it keeps the complexity linear.

Algorithm 2: Envelope Perfect Sampling Algorithm (EPSA)

Data: I.i.d. events $\{u_{-n}\}_{n \in \mathbb{N}} \in \mathcal{A}^{\mathbb{N}}$; the (pre-computed) envelope operation \boxdot

Result: A state $x^* \in \mathcal{S}$ generated according to the stationary distribution of the Markov chain

begin

```

     $n = 1; m := \perp; M := \top;$ 
    repeat
      for  $i = n - 1$  downto 0 do
         $[m, M] := [m, M] \boxdot u_{-i};$ 
       $n := 2n;$ 
    until  $m = M;$ 
     $x^* := m;$ 
    return  $x^*;$ 

```

EPSA is a generalization of both monotone and anti-monotone sampling algorithm (see [14, 8]):

- Monotone events: if an event $a \in \mathcal{A}$ is monotone, then for any $m \preceq M$,

$$[m, M] \boxdot a = [m \cdot a, M \cdot a]. \quad (3.2)$$

- Anti-monotone events: if an event $a \in \mathcal{A}$ is anti-monotone, then for any $m \preceq M$, $[m, M] \boxdot a = [M \cdot a, m \cdot a]$.

In the next section we discuss the complexity of EPSA, that depends on the coupling time of the envelope chain, but also on the complexity of the computation of the envelope transition. The efficiency of the (anti-) monotone case is due to the fact that when ‘ \cdot ’ is (anti-)monotone, the complexity of \boxdot is the same as using ‘ \cdot ’ twice.

Note however that the construction of the envelopes depends on the representation of the Markov chain by a Markov automaton that is not unique. Different automata representing the same chain would lead to different envelope chains with different coupling properties (one may coalesce almost surely and the other not, or if they both coalesce their coupling times may be different). The complexity of the envelope transition may also differ depending on the representation chosen.

3.2 Complexity of EPSA

The envelope approach may not gain over the general PSA because of two drawbacks:

- (D₁) The assumption that $[m_n, M_n]$ hits the set of single point intervals \mathcal{X} in Theorem 3.2 may not hold.
- (D₂) Even if the assumptions for Theorem 3.2 are satisfied, the complexity of the algorithm may be prohibitive. The average complexity of EPSA is $O(\mathcal{C}_e \times \mathbb{E}(\tau_e))$ where τ_e is the number of iterations of the main loop of EPSA (called the coupling time in the following) and \mathcal{C}_e is the complexity (number of elementary operations) of computing $[m, M] \boxdot a$. Meanwhile, the average complexity of the classical PSA is $O(\mathcal{C} \times |\mathcal{S}| \times \mathbb{E}(\tau))$, where τ is the coupling time of PSA, $|\mathcal{S}|$ the cardinality of the state space and \mathcal{C} the complexity of the computation of $x \cdot a$.

Remark 3.3. *It is important to notice that the operation \boxdot can be replaced by any over-approximation \odot such that for all interval $[m, M]$, $[m, M] \boxdot a \subseteq [m, M] \odot a$, without altering the correctness of the algorithm. By doing so, we would still obtain a bounding interval chain, however its coupling time might be longer than for the envelope chain, or the bounding interval chain might not couple at all. There is then an obvious tradeoff between computing simple large bounds quickly with an increased coupling time, or computing tight bounds slowly with a potentially shorter coupling time.*

Remark 3.4. *Actually, even if the envelope chain does not couple, it is still possible to obtain perfect samples in finite time by using the splitting algorithm [2]. The splitting algorithm is hybrid: it first runs EPSA when the envelopes are too far apart and switches to the usual PSA algorithm as soon as the number of states inside the envelopes becomes manageable.*

The rest of this paper focuses on improving the cost \mathcal{C}_e of computing envelopes under several assumptions except for Section 4.2 and numerical examples in Section 7 that discuss the coupling time issue.

Indeed, if \mathcal{C}_e is not too large compared to \mathcal{C} (as it is the case for the various transitions proposed in this article), then the comparison of the two methods is essentially a comparison between the coupling time of EPSA ($\mathbb{E}(\tau_e)$) and the coupling time of PSA multiplied by the cardinality of the state space ($|\mathcal{S}| \times \mathbb{E}(\tau)$).

4 Almost Space Homogeneous Events (ASHEs)

We now present a family of simple events called *Almost Space Homogeneous Events (ASHEs)* for which the corresponding envelope transitions are easy to compute. These events will be the first brick to build more elaborate ones. Intuitively, these events consist in moving homogeneously every point of the space in the same direction, with some simple conditions on the boundaries of the state space.

4.1 Computing the envelopes for ASHEs

We assume from now on that \mathcal{S} is an interval of \mathbb{Z}^d : $\mathcal{S} = [0, C]$, where $C = (C_1, \dots, C_d) \in \mathbb{Z}^d$ endowed by the usual product partial order: $x \leq y \Leftrightarrow x_i \leq y_i, \forall i$. Therefore for any $m, M \in \mathcal{S}$, we have $[m, M] = [m_1, M_1] \times \dots \times [m_d, M_d]$.

Definition 4.1. *An event a is an almost space homogeneous event (ASHE) if there exists a vector $v \in \mathbb{Z}^n$ and a binary relation \mathcal{R} (called blocking relation) on the set of components $\{1, \dots, d\}$ such that for all x in \mathcal{S} we have the following property:*

(P) *Let $CR(x) \stackrel{\text{def}}{=} \{i : x_i + v_i \notin [0, C_i]\}$ be the set of the critical components (i.e. where $x + v$ is out of the state space), and $B(x) \stackrel{\text{def}}{=} \{i : \exists j \in CR(x), (j, i) \in \mathcal{R}\}$ the set of all blocked components in state x . Then for all i :*

$$(x \cdot a)_i = \begin{cases} x_i, & i \in B(x), \\ (x_i + v_i) \vee 0 \wedge C_i, & i \notin B(x). \end{cases} \quad (4.1)$$

In particular, when $x + v \in \mathcal{S}$, we have $x \cdot a = x + v$.

We call v the direction vector of event a .

Remark 4.2. *The blocking relation \mathcal{R} needs only to be specified on $Act \stackrel{\text{def}}{=} \{i \mid v_i \neq 0\}$ the set of active components of the transition, since the event has an effect only along these components.*

The cases $B(x) = \emptyset$ or $B(x) = d$ are interesting particular cases that correspond respectively to the case where, if $x + v \notin \mathcal{S}$, then the event is truncated at the boundary of the state space: $x \cdot a = (x + v) \vee 0 \wedge C$, or the event is blocked: $x \cdot a = x$.

Example 4.3. *Here is a more complex case in the context of queues. Let us consider a Markovian queueing network made of 3 finite queues and exponential service times and a fork node: one departing packet from queue 1 is duplicated into two packets joining queues 2 and 3. This event is an ASHE, corresponds to adding a vector $v = (-1, +1, +1)$ to all states except on the boundaries. The critical cases are when the first queue is empty or when one of the next queues is full. If the first queue is empty, then the event is blocked so that $(1, 2)$ and $(1, 3)$ are in \mathcal{R} . If either queue 2 or queue 3 are full, then the packet actually leaves queue 1 and does not join the other queues (both duplicates are lost). This means that both $(2, 3), (3, 2)$ are in \mathcal{R} . These are the only constraints, thus $\mathcal{R} = \{(1, 2), (1, 3), (2, 3), (3, 2)\}$. Other cases could also be taken into account, for example, the case where a full queue does not prevent the arrival in the other queue, corresponding to $\mathcal{R} = \{(1, 2), (1, 3)\}$, or when a full queue blocks the departure from the first queue, corresponding to $\mathcal{R} = \{(1, 2), (1, 3), (2, 1), (2, 3), (3, 1), (3, 2)\}$.*

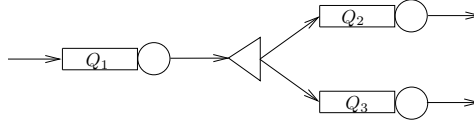


Figure 3: Fork node

The computation of envelopes for a fixed ASHE a is given in Algorithm 3.

Theorem 4.4. *Algorithm 3 computes $[m', M'] = [m, M] \boxplus a$ in $O(d^2)$ elementary operations.*

The proof is given in Appendix A.

4.2 Nonexpansive properties of ASHEs

In the previous Section, we have shown that computing the envelopes for ASHEs can be done efficiently, so that the term \mathcal{C}_e in the complexity is in $O(d^2)$. This section gives some insights on the coupling time. Since our approach is general, no tight bound on the coupling can be derived. The most immediate property is that, as mentioned in Section 3, the coupling time of envelopes is larger than the coupling time of the original chain. However, as seen in the experimental section 7, the difference between two coupling times is not large. This can be explained in part by several nonexpansive properties of ASHEs. In this section we show that the size of the envelope

Algorithm 3: Envelope computation for an ASHE $a \in \mathcal{A}$

Data: States m and M (current infimum and supremum state);

An ASHE a with direction vector v and blocking relation \mathcal{R} ;

Result: $[m', M'] = [m, M] \boxplus a$ (lower and upper envelope).

begin

$Act = \{i : v_i \neq 0\};$

$CR(m) = \{i \in Act : m_i + v_i \notin [0, C_i]\};$

$CR(M) = \{i \in Act : M_i + v_i \notin [0, C_i]\};$

$X = \{j \in Act : \exists i \in CR(m) \cup CR(M) \text{ such that } (i, j) \in \mathcal{R}\};$

$Y = \{j \in Act : \exists i \in CR(m) \cap CR(M) \text{ such that } (i, j) \in \mathcal{R}\};$

 // Note that $Y \subset X$

for $j \in Act \setminus X$ **do**

 // This is the case $j \notin B(m) \cup B(M)$

$m'_j = (m_j + v_j) \vee 0 \wedge C_j; M'_j = (M_j + v_j) \vee 0 \wedge C_j;$

for $j \in Y$ **do**

$m'_j = m_j; M'_j = M_j;$

for $j \in X \setminus Y$ **do**

if $\exists i, i \neq j$ such that $(i, j) \in \mathcal{R}$ and $i \in CR(m) \cup CR(M)$

then

 // component j is blocked by a component i
 different from j

if $v_j < 0$ **then**

$m'_j = (m_j + v_j) \vee 0; M'_j = M_j;$

else

$m'_j = m_j; M'_j = (M_j + v_j) \wedge C_j;$

else

 // only j is blocking j

if $v_j < 0$ **then**

$m'_j = 0; M'_j = (M_j + v_j) \vee (-v_j - 1);$

else

$m'_j = (m_j + v_j) \wedge (C_j - v_j + 1); M'_j = C_j;$

for $j \notin Act$ **do**

$m'_j = m_j; M'_j = M_j;$

interval cannot increase more than linearly in the worse case, and does not increase (or even contracts) under some conditions on the blocking relation. In the latter case, the contracting property also gives some insight of the splitting condition.

Assume that $a \in \mathcal{A}$ is an ASHE given by (4.1). We first give a useful lemma about the set of blocked components of an ASHE.

Lemma 4.5. *Let U be a subset of \mathcal{S} and let $[m, M] = \llbracket U \rrbracket$. We have*

$$\bigcup_{x \in U} B(x) = B(m) \cup B(M).$$

Proof. We first prove the inclusion \subseteq : Let $x \in U$ and $j \in B$. Then there is $i \in CR(x)$ such that $(i, j) \in \mathcal{R}$. Now $i \in CR(x)$ implies that $i \in CR(m) \cup CR(M)$ (as $m_i + v_i \leq x_i + v_i \leq M_i + v_i$), thus $j \in B(m) \cup B(M)$. We now prove the other inclusion. Let $j \in B(m)$, there is a i such that $(i, j) \in \mathcal{R}$ and $m_i + v_i \notin [0, C_i]$. Since $[m, M] = \llbracket U \rrbracket$, there exist $x \in U$ such that $x_i = m_i$, and thus $x_i + v_i \notin [0, C_i]$ and $j \in B(x)$. This shows $B(m) \subseteq \bigcup_{x \in U} B(x)$, and the same reasoning shows $B(M) \subseteq \bigcup_{x \in U} B(x)$. \square

The following proposition shows the nonexpansive property in the case without blocking. For $m, M \in \mathcal{S}$ such that $m \leq M$, denote by $[m', M'] = [m, M] \boxplus a$.

Proposition 4.6. *If for $m, M \in \mathcal{S}$ such that $m \leq M$, $B(m) \cup B(M) = \emptyset$, then:*

$$\|M' - m'\|_1 \leq \|M - m\|_1.$$

Proof. If $B(m) \cup B(M) = \emptyset$, then by Lemma 4.5, $\bigcup_{x \in [m, M]} B(x) = B(m) \cup B(M) = \emptyset$. Then (4.1) becomes $x \cdot a = (x + v) \vee 0 \wedge C$, $\forall m \leq x \leq M$, that is a monotone function of x (i.e. event a is monotone). Therefore, $\|M' - m'\|_1 = \|M \cdot a - m \cdot a\|_1 \leq \|M - m\|_1$. \square

Corollary 4.7. *If $\mathcal{R} = \emptyset$, then for any $m, M \in \mathcal{S}$ such that $m \leq M$:*

$$\|M' - m'\|_1 \leq \|M - m\|_1.$$

The next proposition gives an upper bound in the case where each component can only be blocked by itself. In that case, relation \mathcal{R} is defined by the set \mathcal{B} of all auto-blocking components: $\mathcal{R} = \{(i, i) : i \in \mathcal{B}\}$.

Proposition 4.8. *If $\mathcal{R} = \{(i, i) : i \in \mathcal{B}\}$, for some subset $\mathcal{B} \subset \{1, \dots, d\}$, then:*

$$\begin{aligned} \|M' - m'\|_1 &\leq \sum_{i \notin \mathcal{B}} (M_i - m_i) + \sum_{i \in \mathcal{B}} \max\{M_i - m_i, |v_i| - 1\} \\ &= \|M - m\|_1 + \sum_{i \in \mathcal{B}} [|v_i| - M_i + m_i - 1] \vee 0. \end{aligned}$$

Proof. Since there are no interactions between components, we can consider them separately. Thus, for $i \notin \mathcal{B}$: $M'_i - m'_i \leq M_i - m_i$, as in Proposition 4.6. For $i \in \mathcal{B}$, we have three different cases:

- For all $x \in [m, M]$, $x_i + v_i \in [0, C_i]$. This is equivalent to $i \notin CR(m) \cup CR(M)$. Then $(x \cdot a)_i = x_i + v_i$, $\forall x \in [m, M]$ and $M'_i - m'_i = M_i - m_i$.
- For all $x \in [m, M]$, $x_i + v_i \notin [0, C_i]$. This is equivalent to $i \in CR(m) \cap CR(M)$. Thus $(x \cdot a)_i = x_i$, $\forall x \in [m, M]$ and $M'_i - m'_i = M_i - m_i$.
- There are $x, y \in [m, M]$ such that $x_i + v_i \in [0, C_i]$ and $y_i + v_i \notin [0, C_i]$. Suppose that $v_i > 0$ (the case $v_i < 0$ is symmetrical). Then $i \in CR(M) \setminus CR(m)$. Denote by $x'_i = C_i - v_i$. Then $m_i \leq x'_i < M_i$. We have:

$$m'_i = (m_i + v_i) \wedge (x'_i + 1), \quad M'_i = C_i.$$

There are two cases:

- $m_i + v_i \leq x'_i + 1$. Then $M'_i - m'_i = C_i - (m_i + v_i) = x'_i - m_i \leq M_i - m_i - 1$.
- $m_i + v_i > x'_i + 1$. Then $M'_i - m'_i = C_i - (x'_i + 1) = v_i - 1$.

Thus $M'_i - m'_i \leq \max\{M_i - m_i, |v_i| - 1\}$. \square

The above result tells us that an event without blocking interactions between different components (*i.e.* $(i, j) \in \mathcal{R} \implies i = j$) is nonexpansive in components that are not auto-blocking. For components that are auto-blocking such event is nonexpansive when $M_i - m_i \geq |v_i| - 1$.

Corollary 4.9. *If $\mathcal{R} = \{(i, i) : i \in \mathcal{B}\}$, for some subset $\mathcal{B} \subset \{1, \dots, d\}$, and $|v_i| = 1, \forall i \in \mathcal{B}$, then $\|M' - m'\|_1 \leq \|M - m\|_1$, $m \leq M$.*

We give now an upper valid bound for any ASHE.

Proposition 4.10. *For any ASHE 'a' given by (4.1), we have:*

$$\|M' - m'\|_1 \leq \|M - m\|_1 + \|v\|_1 - 1.$$

Proof. For $i \in \{1, \dots, d\}$ such that $v_i < 0$: $m_i + v_i \leq (x \cdot a)_i \leq M_i$, $\forall x \in [m, M]$, and for $i \in \{1, \dots, d\}$ such that $v_i > 0$: $m_i \leq (x \cdot a)_i \leq M_i + v_i$, $\forall x \in [m, M]$. Thus $\|M' - m'\|_1 \leq \|M - m\|_1 + \|v\|_1$.

Suppose that for at least one component $i \in A$, we have two states x and y such that $x_i + v_i \in [0, C_i]$ and $y_i + v_i \notin [0, C_i]$. (Otherwise $\|M' - m'\|_1 \leq \|M - m\|_1$.) For this component: $(M'_i - m'_i \leq M_i - m_i + |v_i| - 1$. Indeed, suppose that $v_i > 0$ (the case $v_i < 0$ is symmetrical). Then $M_i + v_i - 1 \geq C_i$ and $m_i \leq (z \cdot a)_i \leq C_i, \forall z \in [m, M]$. Thus $M'_i - m'_i \leq C_i - m_i \leq M_i - 1 + v_i - m_i$, so we have:

$$\|M' - m'\|_1 \leq \|M - m\|_1 + \|v\|_1 - 1.$$

\square

5 Partition of the state space

5.1 Zone partition

In this section we focus on a single event a . The state space is *partitioned into zones*, Z^1, \dots, Z^K , where K is assumed to be small compared to the size of the state space. The action a may have a different effect a^k on each zone Z^k . Formally, the action ‘ \cdot ’ by a is defined for all $x \in \mathcal{S}$ as : $x \cdot a = x \cdot a^k$ where k is such that $x \in Z^k$. To compute the envelopes of the action of a over an arbitrary interval $[m, M]$, one can use the partition into zones.

$$\begin{aligned}
 [m, M] \sqsupset a &= \llbracket [m, M] \cdot a \rrbracket \\
 &= \llbracket \left(\bigcup_{k=1}^K [m, M] \cap Z^k \right) \cdot a \rrbracket \\
 &= \llbracket \bigcup_{k=1}^K ([m, M] \cap Z^k \cdot a) \rrbracket \\
 &= \llbracket \bigcup_{k=1}^K ([m, M] \cap Z^k \cdot a^k) \rrbracket \\
 &= \llbracket \bigcup_{k=1}^K \llbracket [m, M] \cap Z^k \rrbracket \cdot a^k \rrbracket \tag{5.1} \\
 &\subset \llbracket \bigcup_{k=1}^K \llbracket [m, M] \cap Z^k \rrbracket \sqsupset a^k \rrbracket . \tag{5.2}
 \end{aligned}$$

The construction of these envelopes is illustrated in Figure 4. The first and second equalities are direct consequence of the definition of the partition; the third one is by definition of the action on a set; the fourth one simply uses the fact that $\llbracket \llbracket A \rrbracket \cup \llbracket B \rrbracket \rrbracket = \llbracket A \cup B \rrbracket$.

The last inclusion (5.2) does not hold with equality in general. It uses the fact that for any set A , $\llbracket A \cdot a \rrbracket \subset \llbracket A \rrbracket \sqsupset a$. It gives us an easy way to compute an over-approximation of \sqsupset . Moreover, in some cases the inclusion can be replaced by an equality, such that the over-approximation is exact (or *tight*). We give some simple cases where this equality holds in Section 5.2.

The computation of an interval containing $[m, M] \sqsupset a$ using (5.2) reduces to two problems:

- (P_1) The computation of $\llbracket [m, M] \cap Z^k \rrbracket$. This can be done by integer linear programming (see next section).
- (P_2) The computation of $\llbracket [m, M] \cap Z^k \rrbracket \sqsupset a^k$. This is easy when a^k is an ASHE (using Algorithm 3) or when a^k is monotone (see Equation

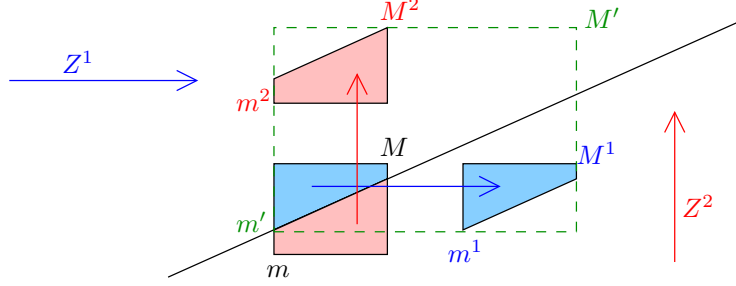


Figure 4: The construction of envelopes on a simple example. The initial interval $[m, M]$ intersects the two zones Z^1 and Z^2 . We compute the next interval $[m', M'] \stackrel{\text{def}}{=} [m, M] \boxdot a$ by computing $[m^1, M^1] = \llbracket [m, M] \cap Z^1 \cdot a \rrbracket$, $[m^2, M^2] = \llbracket [m, M] \cap Z^2 \cdot a \rrbracket$, and finally $[m', M'] = \llbracket [m^1, M^1] \cup [m^2, M^2] \rrbracket$ correspond to (5.1), but in this example the equality holds also in the inclusion (5.2): $[m', M'] = \llbracket \llbracket [m, M] \cap Z^1 \rrbracket \boxdot a \cup \llbracket [m, M] \cap Z^2 \rrbracket \boxdot a \rrbracket$.

(3.2)).

In Section 6, we will give methods to solve (P_1) by computing $\llbracket [m, M] \cap Z^k \rrbracket$ or an over-approximation.

5.2 Tightness

Proposition 5.1. *Let U be a subset of \mathcal{S} and a an ASHE. If a is non-blocking on U (i.e. $\forall x \in U, B(x) = \emptyset$), then $\llbracket U \cdot a \rrbracket = \llbracket U \rrbracket \boxdot a$.*

Proof. The inclusion $\llbracket U \cdot a \rrbracket \subseteq \llbracket U \rrbracket \boxdot a$ is simple, since by definition $\llbracket U \rrbracket \boxdot a = \llbracket \llbracket U \rrbracket \cdot a \rrbracket$ and $U \subseteq \llbracket U \rrbracket$.

Let us prove now the other inclusion. Let $[m, M] = \llbracket U \rrbracket$. Since we assumed a to be non-blocking on U , i.e. $\bigcup_{x \in U} B(x) = \emptyset$, we have for all x in U , $x \cdot a = (x + v) \vee 0 \wedge C$. Using Lemma 4.5, we also have $B(m) \cup B(M) = \bigcup_{x \in U} B(x) = \emptyset$, which gives $[m, M] \boxdot a = [(m + v) \vee 0 \wedge C, (M + v) \vee 0 \wedge C]$ (see Algorithm 3, in which $B(m) \cup B(M) = \emptyset$ implies that $X = \emptyset$).

We can now prove that $\llbracket U \rrbracket \boxdot a \subseteq \llbracket U \cdot a \rrbracket$. Let $[m', M'] = \llbracket U \rrbracket \boxdot a$ and $[q, Q] = \llbracket U \cdot a \rrbracket$. To prove the inclusion we just need to show that $m' \geq q$ and $M' \leq Q$.

We showed that $m' = (m + v) \vee 0 \wedge C$. For each component $i \in \{1, \dots, d\}$, there exists $x \in U$ such that $x_i = m_i$. Let $x' = (x + v) \vee 0 \wedge C$, we have $x' \in U \cdot a$, thus $x' \geq q$, and $m'_i = m_i + v_i \vee 0 \wedge C_i = x_i + v_i \vee 0 \wedge C_i = x'_i \geq q_i$. Finally, $m' \geq q$, and similarly $M' \leq Q$. \square

Corollary 5.2. *Let a be a piecewise ASHE over a partition Z^1, \dots, Z^K such that each ASHE a^k is non-blocking on its zone Z^k . Then the inclusion (5.2) can be replaced by an equality.*

Remark 5.3. *It is easy to see that any ASHE a can be partitioned into zones such that a is non-blocking on each zone. Although this is not useful for computing the envelope of a simple ASHE (since we already know how to compute it), it can however be used to obtain tight envelopes on piecewise ASHEs.*

Proposition 5.4. *Let U be a subset of \mathcal{S} , a an ASHE and v its direction vector. If a is unidimensional, (i.e. $\text{Act} = \{i \mid v_i \neq 0\}$ is a singleton), and if the projection of U on any component i : $U_i \stackrel{\text{def}}{=} \{x_i \mid x \in U\}$ is an interval, then $\llbracket U \cdot a \rrbracket = \llbracket U \rrbracket \boxplus a$.*

Proof. The inclusion $\llbracket U \cdot a \rrbracket \subseteq \llbracket U \rrbracket \boxplus a$ is always true, and was already proved. We keep the same notation as in the previous proof: $[m, M] = \llbracket U \rrbracket$, $[m', M'] = \llbracket U \rrbracket \boxplus a$ and $[q, Q] = \llbracket U \cdot a \rrbracket$. Let i be the component where $v_i \neq 0$. For all $j \neq i$, we have $m'_j = m_j$, $M'_j = M_j$, and for all $x \in U$, $(x \cdot a)_j = x_j$, thus $m'_j = q_j$ and $M'_j = Q_j$ clearly.

On component i , we need to consider several cases (we will show that $m'_i \geq q_i$ and $M'_i \leq Q_i$):

- If $i \notin B(m) \cup B(M)$, then the transition is non-blocking on U , and the previous Proposition applies.
- If $i \in B(m) \cap B(M)$, then $[m', M'] = [m, M]$ and $U \cdot a = U$, thus $[m', M'] = [q, Q]$.
- If $i \in B(m) \setminus B(M)$, then $m_i < -v_i < M_i$. In that case, we have $m'_i = 0$ and $M'_i = (M_i + v_i) \vee (-v_i - 1)$. By assumption, $U_i = \{x_i \mid x \in U\}$ is an interval, and thus $U_i = [m_i, M_i]$. Since $m_i < -v_i < M_i$, there exists $x \in U$ such that $x_i = -v_i$. We then have $(x \cdot a)_i = x_i + v_i = 0$, and thus $q_i = 0 = m'_i$.

We now prove that $Q_i \geq M'_i = (M_i + v_i) \vee (-v_i - 1)$. Since $U_i = [m_i, M_i]$, there exists $y \in U$ such that $y_i = M_i$, thus $Q_i \geq (y \cdot a)_i = y_i + v_i = M_i + v_i$. Also, since $m_i \leq -v_i - 1 < M_i$, there exists a $z \in U$ such that $z_i = -v_i - 1$. We then have $z_i + v_i = -1 < 0$, and thus $i \in B(z)$ and $(z \cdot a)_i = z_i = -v_i - 1$, and finally $Q_i \geq (z \cdot a)_i \geq -v_i - 1$.

- If $i \in B(M) \setminus B(m)$ a similar reasoning shows the same result.

We finally obtain the reverse inclusion $\llbracket U \rrbracket \boxplus a \subseteq \llbracket U \cdot a \rrbracket$.

□

6 Polytopic Zones and Linear Programming

In this section we consider a class of piecewise events for which the computation of envelopes can be done in a reasonable amount of time and memory

space. We will focus here on one event a assumed to be defined piecewise over polytopic zones.

Again, the state space is a product $\mathcal{S} = [0, C_1] \times \cdots \times [0, C_d]$, and is here partitioned into zones defined by linear constraints.

Let us consider H hyperplanes of \mathbb{R}^d in general position and assume they do not intersect \mathcal{S} (by translating them by a small ε if needed). These hyperplanes define a partition of \mathbb{R}^d in K convex polytopes (P^1, \dots, P^K) , and these polytopes induce a partition of the state space \mathcal{S} in K zones $Z^i = \mathcal{S} \cap P^i$. Without loss of generality, we consider only the nonempty zones and assume we have K of them.

Finally, over each zone Z^k , the event a^k is assumed to be easily computable (e.g. an ASHE).

There exists a relation between the total number of hyperplanes H needed to define the zones and the maximal number of zones that can be defined using them. The number K of zones is smaller than $1 + H + \binom{H}{2} + \cdots + \binom{H}{d}$. This is because the state space \mathcal{S} may intersect all zones (bounded or not) formed by H hyperplanes in a general position. The number of such zones is given by the value of the characteristic polynomial of the hyperplanes at point -1 (see [17]).

Example 6.1. *A queueing example is displayed in Figure 5. We consider a Join the Shortest Waiting time (JSW) event, that is an arrival that allocates the new customer to the queue with the shortest expected waiting time (after arrival). The zones for the JSW event are displayed in Figure 6. If $(x_2 + 1)/\mu_2 > (x_1 + 1)/(2\mu_1) + \varepsilon$ (with ε very small) then the new customer is sent to queue 1 as long as the queue is not full, in which case it is sent to queue 2. If $(x_2 + 1)/\mu_2 \leq (x_1 + 1)/(2\mu_1) + \varepsilon$, it is sent to queue 2 as long as the queue is not full, in which case it is sent to queue 1. If both queues are full, the client is rejected. Here 3 hyperplanes divide the two-dimensional state space into 4 zones that are polytopes: $(x_2 + 1)/\mu_2 = (x_1 + 1)/(2\mu_1) + \varepsilon$, $x_1 = C_1 - 1 + \varepsilon$, $x_2 = C_2 - 1 + \varepsilon$.*

Note that in this case, the partition makes the event a non-blocking ASHE in each zone, so that Equation (5.2) holds with equality (by Proposition 5.1). However, the two zones in the upper right corner could be merged in one single zone with a blocking ASHE, and the equality would still hold (see Proposition 5.4).

6.1 Zone intersection with Linear Programming

Since we assumed the event a to be easily computable on each zone Z^k , the only problem left to solve is (P_1) the computation of $[m', M'] = \llbracket [m, M] \cap Z^k \rrbracket$. When the zones Z^k are induced by polytopes P^k ($Z^k = \mathcal{S} \cap P^k$), this computation amounts to solving $2d$ integer linear programs, one for each component m'_i and M'_i :

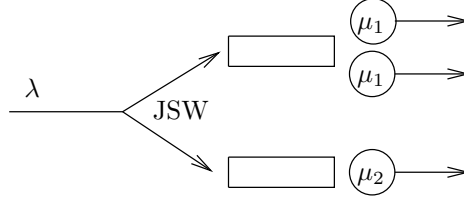


Figure 5: Example of a queueing event that can be defined with ASHEs on polytopic zones. The first queue has two identical servers with rate μ_1 while the second has only one server with rate μ_2 .

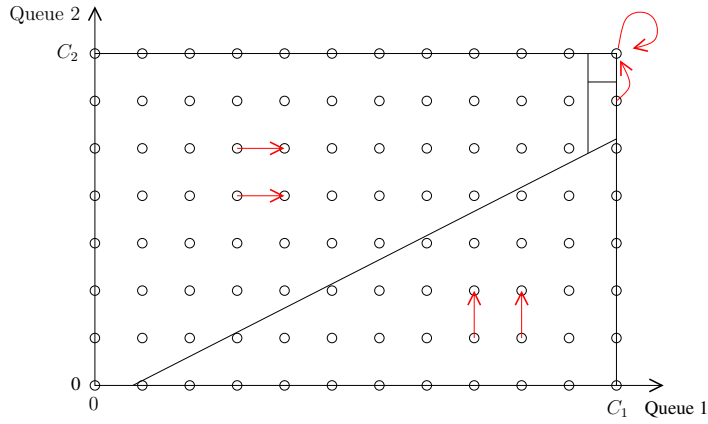


Figure 6: The zones for the JSW event when $(2\mu_1 > \mu_2)$.

$$\forall i \in \{1, \dots, d\}, \quad \begin{cases} m'_i &= \min\{x_i \mid x \in \mathbb{Z}^d \cap P^k \cap [m, M]\}, \\ M'_i &= \max\{x_i \mid x \in \mathbb{Z}^d \cap P^k \cap [m, M]\}. \end{cases} \quad (6.1)$$

Unfortunately, integer linear programs are well known to be NP-hard in general. These linear programs can be solved in simple cases that we will not detail in this paper. A general approach to overcome this difficulty is to relax the integrity condition and to solve the corresponding rational linear programs, by computing q^k and Q^k in \mathbb{Z}^d defined by

$$\forall i \in \{1, \dots, d\}, \quad \begin{cases} q_i^k &\stackrel{\text{def}}{=} \lceil \min\{x_i \mid x \in \mathbb{Q}^d \cap P^k \cap [m, M]_{\mathbb{Q}}\} \rceil, \\ Q_i^k &\stackrel{\text{def}}{=} \lfloor \max\{x_i \mid x \in \mathbb{Q}^d \cap P^k \cap [m, M]_{\mathbb{Q}}\} \rfloor, \end{cases} \quad (6.2)$$

where $[m, M]_{\mathbb{Q}}$ denotes the rational interval $\{x \in \mathbb{Q}^d \mid m \leq x \leq M\}$, that contains $[m, M]$.

It is easy to see that the discrete interval $[q^k, Q^k]$ contains $\llbracket [m, M] \cap \mathbb{Z}^k \rrbracket$. We can thus compute an over-approximation of \square for a :

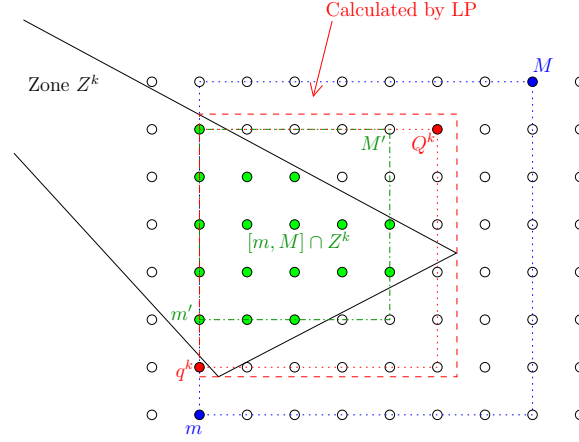


Figure 7: Consider a zone Z^k and an interval $[m, M]$. The Figure displays the intervals $[m, M]$, $[m', M'] \stackrel{\text{def}}{=} \llbracket [m, M] \cap Z^k \rrbracket$ and the larger interval $[q^k, Q^k]$ constructed using (6.2).

$$\begin{aligned}
 [m, M] \sqcup a &\subset \left[\bigcup_{k=1}^K \llbracket [m, M] \cap Z^k \rrbracket \sqcup a^k \right] \\
 &\subset \left[\bigcup_{k=1}^K [q^k, Q^k] \sqcup a^k \right].
 \end{aligned}$$

Figure 7 gives an illustration on a simple example.

6.2 Complexity and data structure

In order to deal with the set of linear programs in an efficient way, they can be represented as a binary acyclic graph. Each node corresponds to an additional linear constraint. The left child corresponds to the half plane where the inequality is negative and the right child to the one where the inequality is positive.

When an intersection can be reached through several paths, the nodes are merged.

The leaves of the graph are all the polytopes and the number of nodes in the tree is less than twice the number of leaves. This construction is illustrated in Figure 8.

This representation has a better amortized complexity to compute $[m, M] \cdot a$ than an exhaustive resolution for each polytopic zone since the intersection of $[m, M]$ with a zone can be found empty early in the tree (this can even happen at the root of the tree if $[m, M]$ does not intersect one half plane).

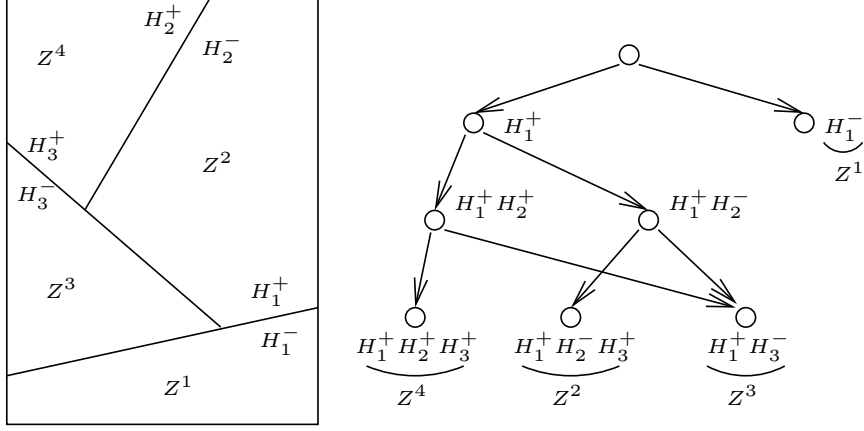


Figure 8: A partition and the corresponding graph structure used to sweep the zones. Here, the space is partitioned into 4 polytopic zones, Z_1, Z_2, Z_3 and Z_4 , defined by three hyperplanes H_1, H_2 and H_3 . The corresponding half-spaces are denoted H_i^+ and H_i^- , for all i . Hence, for instance, $Z^2 = H_1^+ \cap H_2^- \cap H_3^+$.

Sorting the constraints (from the root of the tree to the leaves) to get the best possible expected cost of computing $[m, M] \cdot a$ can be done for a given distribution of m and M (for example uniform over $\mathcal{S} \times \mathcal{S}$) by using dynamic programming. This is not in the scope of this paper.

The Algorithm 4 defines the procedure $\text{Bound}(V, m, M, a)$ such that, when V is the graph G , $\text{Bound}(G, m, M, a) = \left\llbracket \bigcup_{k=1}^K [q^k, Q^k] \sqcap a^k \right\rrbracket$ which is an over-approximation of $[m, M] \sqcap a$. It uses the procedure LP-Solve (not detailed) that solves linear programs over rational numbers in polynomial time. It also contains the test $V \cap [m, M]_{\mathbb{Q}} \neq \emptyset$ that can be done much faster (in linear time, see the next section) and helps to cut the cost: The LP-solve procedure is not used when the intersection is empty.

By solving linear programs with the interior point method [10], the complexity of each main call to Algorithm 4 is $O(d^{4.5}K)$.

6.3 Fast bounds using Minkowski sums

The previous cost to compute over-approximations of envelopes by solving linear programs may not be sustainable for chains with a large dimension d and when the number of calls to *Bound* is large.

In this section we will derive a way to compute looser bounds on the envelopes, with a complexity $O(dHK)$. This technique is based on the computation of Minkowski sums of all K zones with intervals $[m, M]$ (see [13]).

Each polytope P^k can be expressed as the intersection of h_k hyperplanes: $P^k = \{x \in \mathbb{R}^d \mid xA^k \leq b^k\}$, where A^k is a $d \times h_k$ matrix and b^k a row vector of

Algorithm 4: Recursive procedure computing the bounds on the envelopes using linear programming

Data: Partition of the state space into Z^1, \dots, Z^K arranged in a graph structure G using linear constraints;

Result: interval $[m', M'] = \left[\bigcup_{P^k \cap [m, M]_{\mathbb{Q}} \neq \emptyset} [q^k, Q^k] \boxplus a^k \right]$.

Bound $(V, m, M, a) :=$

begin

Visited(V) := True;

if $V \cap [m, M]_{\mathbb{Q}} \neq \emptyset$ **then**

if $V = P^k$ **then**

for $j = 1$ **to** d **do**

$q_j^k := \lceil \text{LP-Solve}(\min\{x_j \mid x \in V \cap [m, M]_{\mathbb{Q}}\}) \rceil$;

$Q_j^k := \lfloor \text{LP-Solve}(\max\{x_j \mid x \in V \cap [m, M]_{\mathbb{Q}}\}) \rfloor$;

$[m', M'] := \llbracket [m', M'] \cup [q^k, Q^k] \boxplus a^k \rrbracket$;

else

if Visited($V.\text{left}$) = False **then**

Bound($V.\text{left}, m, M, a$)

if Visited($V.\text{right}$) = False **then**

Bound($V.\text{right}, m, M, a$)

size h_k . Each constraints $\{A_{(\cdot, j)}^k x \leq b_j^k\}$ correspond to one of the hyperplane defining P^k .

Testing if an interval $[m, M]_{\mathbb{Q}}$ of size $s \stackrel{\text{def}}{=} M - m$ intersects a polytope P^k can be done by testing if the center of the interval $(m + s/2)$ belongs to an extended zone W^k that is the *Minkowski sum* of P^k and the interval $[-s/2, s/2]$: $W^k \stackrel{\text{def}}{=} P^k \oplus [-s/2, s/2]$ (the Minkowski sum $A \oplus B$ of two sets A and B is defined by $A \oplus B = \{a + b \mid a \in A, b \in B\}$).

Figure 9 shows an illustration of the Minkowski sum in a simple case.

It is direct to see that

$$[m, M]_{\mathbb{Q}} \cap P^k \neq \emptyset \Leftrightarrow m + s/2 \in W^k.$$

and that the set W^k is a polytopic zone defined by $h_k + 2d$ hyperplanes:

$$W^k = \{x \in \mathbb{R}^d \mid xA^k \leq b^k + (s/2)|A^k|\} \cap \{x \in \mathbb{R}^d \mid x \geq \ell^k - s/2, x \leq L^k + s/2\},$$

where $|A^k| \stackrel{\text{def}}{=} (|A_{i,j}^k|)_{1 \leq i \leq d, 1 \leq j \leq h_k}$ is the component-wise absolute value of matrix A^k , and ℓ^k and L^k are defined in \mathbb{Q}^d by

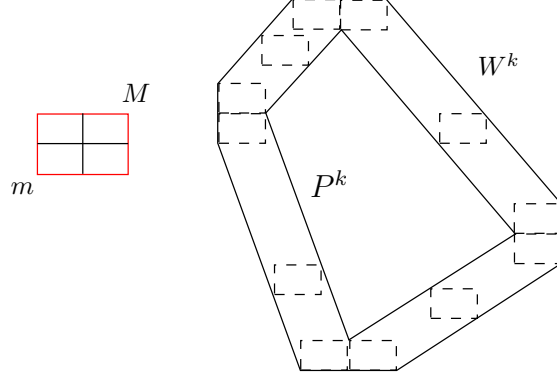


Figure 9: The Minkowski sum of a polytope P^k and the interval $[-s/2, s/2]$, s being the size of the interval $[m, M]$.

$$\forall i \in \{1, \dots, d\}, \quad \begin{cases} \ell_i^k & \stackrel{\text{def}}{=} \min\{x_i \mid x \in \mathbb{Q}^d \cap P^k\}, \\ L_i^k & \stackrel{\text{def}}{=} \max\{x_i \mid x \in \mathbb{Q}^d \cap P^k\}. \end{cases} \quad (6.3)$$

Note that ℓ^k and L^k do not depend on $[m, M]$ and thus can be pre-computed with linear programming before starting the simulation.

Testing if the center of $[m, M]$ belongs to W^k simply amounts to testing if the center satisfies all inequalities defining W^k . This is done in time $O(dh_k)$, and H being the total number of hyperplanes used to define the polytopes, we have $h_k \leq H$. Therefore, the test $[m, M] \cap P^k \neq \emptyset$ in Algorithm 4 can be done in time $O(dH)$ for each polytope P^k .

The previous interval $[q^k, Q^k]$ computed with linear programs (solvable in $O(d^4)$) can be replaced by a looser over-approximation of $\llbracket Z^k \cap [m, M] \rrbracket$ computable in time $O(2d)$.

Indeed, since $P^k \subset [\ell^k, L^k]_{\mathbb{Q}}$, we have

$$\llbracket Z^k \cap [m, M] \rrbracket \subset [\lceil \ell^k \rceil, \lfloor L^k \rfloor] \cap [m, M] = [m \vee \lceil \ell^k \rceil, M \wedge \lfloor L^k \rfloor],$$

where $\lceil \ell^k \rceil$ (resp. $\lfloor L^k \rfloor$) is the component-wise ceiling (resp. floor) of ℓ^k (resp. L^k) and is pre-computed. The intervals $[m \vee \lceil \ell^k \rceil, M \wedge \lfloor L^k \rfloor]$ can replace the intervals $[q^k, Q^k]$ in Algorithm 4, and are easier to compute. The complexity of Algorithm 4 would then be $O(dHK)$. However, these intervals provide looser bounds on the envelopes and might thus increase the coupling time of the bounding processes, at worst to infinity. As we pointed out earlier, this tradeoff between coupling time and complexity of computations needs to be adjusted depending on the Markov chain to simulate.

7 Queueing Network Examples

In this section, we illustrate the usefulness and the expressive power of piecewise ASHES by showing examples of queueing networks that can be described using Markov automata with piecewise ASHES over polytopic zones.

The rule of thumb is that many classical (and some non-classical) queueing networks are such chains so that sampling their stationary distribution is made possible using ASHES over polytopic zones, while the classical PSA would be intractable in practice.

7.1 Jackson Network with finite capacities

We consider open queueing systems similar to Jackson networks except that queues have finite capacities. The network is formed by d queues numbered from 1 to d : Each queue i is a M/M/1/ C_i queue with finite capacity C_i and a FIFO server. Entering customers arrive in each server i according to a Poisson process of parameter λ_i , and each server is assumed to have exponential service time μ_i . After being served at queue i , a customer is directed to queue j with probability $P_{i,j}$ or leaves the network with probability $P_{i,0} \stackrel{\text{def}}{=} 1 - \sum_{j=1}^d P_{i,j}$. We will consider two types of blocking policy:

- Client Loss (CL): If a client served by queue i is sent to a full queue, it is lost.
- Restart Service (RS): If a client served by queue i is sent to a full queue, it restarts its service at queue i .

We assume that new clients arriving in the network from outside are lost when sent to a full queue. The state of the network is entirely described by the number of customers in each queue, and the corresponding Markov chain can be generated with the following Markov automaton $\mathcal{A} = (\mathcal{S}, A, D, \cdot)$ where $\mathcal{S} = [0, C_1] \times \dots \times [0, C_d]$, $A = \{a_{i,j} | 0 \leq i, j \leq d\}$, $D(a_{i,j}) = \lambda_{a_{i,j}} / \sum_{a \in A} \lambda_a$, where λ_a the frequency rate of each event a is defined as:

$$\begin{cases} \lambda_{a_{0,0}} := 0 \\ \lambda_{a_{0,j}} := \lambda_j & \forall j \in [1, \dots, d] \\ \lambda_{a_{i,j}} := \mu_i P_{i,j} & \forall i, j \in [1, \dots, d] \end{cases}$$

and where the $a_{i,j}$ are ASHES with direction vector $v_{a_{i,j}} = e_j - e_i$ for all $i, j \in \{0, \dots, d\}$ (with the convention $e_0 = (0, \dots, 0)$), and with a blocking relation $\mathcal{R}_{a_{i,j}}$ depending on the blocking policy used: $\mathcal{R}_{a_{i,j}}^{CL} = \{(i, j)\}$ for the Client Loss policy and $\mathcal{R}_{a_{i,j}}^{RS} = \{(i, j), (j, i)\}$ for the Service Restart policy. Note that when $i = 0$ or $j = 0$, the choice of \mathcal{R} has no incidence. The blocking policy may also differ for each couple i, j .

It is easy to check that the events of this Markov automaton are all monotone for the product order on $\mathcal{S} = [0, C_1] \times [0, C_d]$. Therefore for this type of networks, the coupling times for PSA and EPSA are equal.

7.2 Examples of non-monotone ASHEs

In this section, we consider other types of routing events that are not monotone nor anti-monotone and can be described with ASHEs. These events can easily be added to any Jackson networks with finite capacity, and they can easily be combined with each other to give other types of ASHEs.

Envelopes for these events were already computed in [2] using ad-hoc methods, while the ASHE framework now embraces them all. The numerical results for these events are taken from [2].

Fork and Join A common type of non-monotone event used in queueing network is fork and join. A fork queue is such that, at service completion, several customers are sent simultaneously to several disjoint output queues. A join is a queue with several input buffers. The server is only active when no buffer is empty. A service completion removes one customer in each input buffer and sends one (merged) customer to an output buffer.

Fork events were already introduced in Example 4.3 and shown to be ASHEs. A join event with two input buffers i and j and output buffer k is described by an ASHE with direction vector $v = e_k - e_i - e_j$ and blocking relation depending on the blocking policy: $\mathcal{R}^{CL} = \{(i, k), (j, k), (i, j), (j, i)\}$ or $\mathcal{R}^{RS} = \{(i, k), (j, k), (i, j), (j, i), (k, i), (k, j)\}$.

Join events are not monotone for the natural product order: Consider for instance the join event a with $i = 1, j = 2$ and $k = 3$ in a 3-server network, and consider the two states $x = (0, 1, 0)$ and $y = (1, 1, 0)$. Then $x < y$ while $x \cdot a = (0, 1, 0)$ and $y \cdot a = (0, 0, 1)$ are incomparable. Similarly, it is easy to show that fork is in general not monotone.

Negative customers We consider the additional events corresponding to negative customers. After queue i finishes a service, a negative customer is sent to queue j and kills one customer in that queue, if there is any. This event can be represented by an ASHE of direction vector $v = -e_i - e_j$ and blocking relation $\mathcal{R} = \{(i, j)\}$. Such events are not monotone for the natural product order.

We ran simulations for the small network displayed in Figure 11. The average coupling time of PSA and EPSA are reported in Figure 10 (in the following all data are given with 95 % confidence intervals).

In the example, the coupling time of EPSA is around 3 times larger than PSA. Since the number of states is $15^2 = 225$, this means that EPSA is about $225/6 = 40$ times faster than PSA in this case.

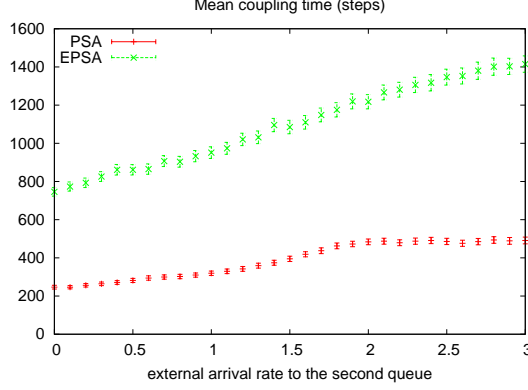


Figure 10: Mean coupling times of PSA and EPSA algorithms for the network in Figure 11 as a function of λ_2 with 95 % confidence intervals.

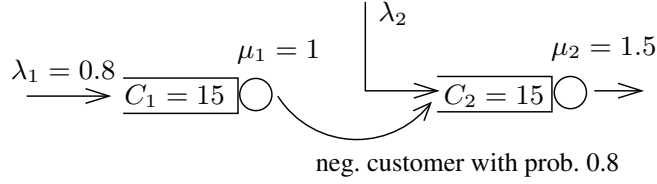


Figure 11: A network with negative customers.

Batch routing Here, we consider batch routing: After the end of service in queue i , a batch of K customers leaves queue i and creates L new customers in queue j . In general, K and L do not need to be equal: for instance 10 products that are assembled into one (*e.g.* put into a box), and then sent to next phase of a production line can be modeled by taking $K = 10$ and $L = 1$. This event can be represented by an ASHE of direction vector $v = -Ke_i + Le_j$ and blocking relation depending on the blocking policy: $\mathcal{R}^{CL} = \{(i, i), (i, j)\}$ or $\mathcal{R}^{RS} = \{(i, i), (i, j), (j, i), (j, j)\}$. As before, the exogenous arrivals (resp. services) can be modeled by setting $i = 0$ (resp. $j = 0$). We allow partial arrivals to queue j (in the case of \mathcal{R}^{CL}), while the service must be complete (*i.e.* at least K customers must be present in queue i). However, partial service can also be modeled by piecewise ASHEs, using similar ideas as for multiple servers in Section 7.3.

Figure 12 reports the coupling time of EPSA for a single queue with batch arrivals of size $K = 1, 2, 3$ or 5 with rate λ , departures of size 1 and rate 1, and blocking policy $\mathcal{R} = \{(1, 1)\}$. Unlike what happens for negative customers and fork-join nodes, the coupling time of EPSA increases exponentially fast with the load of the queue (the scale being logarithmic).

Such increase is due to the fact that envelopes can only couple after they both hit 0. The time for the upper envelope starting in C to hit 0 when the load ρ is larger than 1 becomes exponentially large (larger than $e^{\rho C}$).

One way to prevent this is to use the Split algorithm [2, Algorithm 3], that is an hybrid between PSA and EPSA: It starts by computing bounding intervals as EPSA until the size of the interval becomes smaller than the size of the largest batch. When this happens, the bounding interval is then split into all the single trajectories contained in it, and these normal trajectories are then computed as in PSA. The advantage of the Split algorithm is that these normal trajectories are not as numerous as the $|\mathcal{S}|$ trajectories needed for the classical PSA, and it is guaranteed to terminate when PSA does unlike EPSA.

Figure 13 reports the coupling times of EPSA, PSA and Split for a queue with capacity 20, arrival batches +2 and +3 with probabilities 0.49 and 0.51, and departures of size one (with rate 1). In this figure, the coupling time of PSA and Split are similar while EPSA's coupling time grows very fast when λ/μ is larger than one. In this case Split has a clear advantage over the other two algorithms: It couples exponentially faster than EPSA and deals with two envelopes instead of C trajectories for PSA. The efficiency of Split can be partially explained by the choice of the splitting point: as long as the size of the bounding interval is larger than K (the size of a batch), the batch arrival event is nonexpansive. This is a direct consequence of Proposition 4.8.

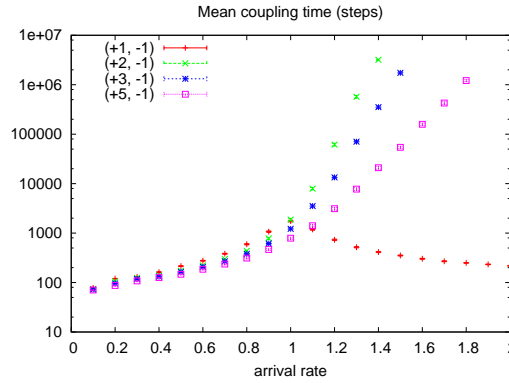


Figure 12: Mean coupling times for EPSA in a $(+k, -1)$ queue with batch arrivals of size $k = 1, 2, 3$ and 5 respectively, as a function of the arrival rate λ .

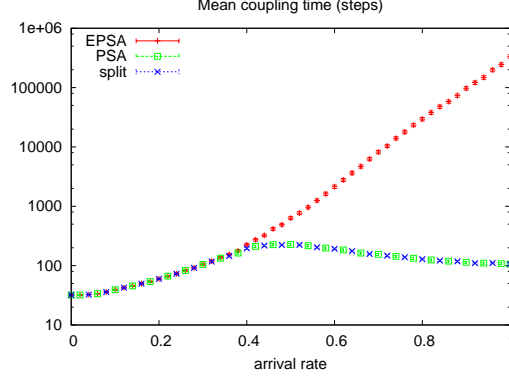


Figure 13: Mean coupling times for PSA, EPSA and Split for a $(+2, +3, -1)$ queue.

7.3 Examples of Piecewise ASHEs

We now consider even more complex queueing events that will prove to be piecewise ASHEs over polytopic zones.

Multiple Servers Let us assume now that queue i is a $M/M/C/C$ queue with finite capacity C and C identical servers with serving rate μ , and that customers are sent to queue j after being served. Such type of service can be described using C piecewise ASHEs a_1, \dots, a_C as illustrated in Figure 14.

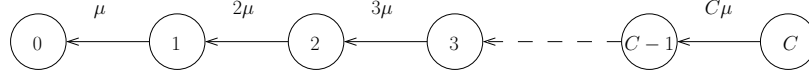
The state space of queue i is $[0, C]$ and the transition rate matrix Q associated to this event is such that $Q(x, x-1) = x\mu$, the other coordinates being null. We add C events a_1, \dots, a_C with each a_k defined as the following piecewise ASHE: on zone $[0, k-1]$, a_k is the identity ($v = 0$, $\mathcal{R} = \emptyset$) and on $[k, C]$, a_k is the ASHE with direction vector ($v = e_j - e_i$) and blocking relation $\mathcal{R} = \emptyset$. Finally, we set $\lambda_{a_k} = \mu$, and take $D(a_k) = \lambda_{a_k} / \sum_{a \in A} \lambda_a$.

This construction can be easily adapted to any queue with $n < C$ servers, which can be represented with n piecewise ASHEs. It can also be easily combined with any other type of events discussed previously.

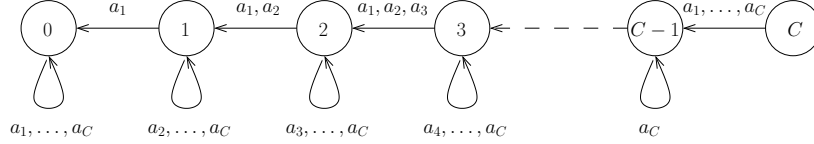
Join the Shortest Waiting Time (JSW) and other routing policies

JSW is described in Example 6.1 and is shown to be a piecewise ASHE over polytopic zones.

More generally, consider the routing event r_i where a client, after being served at queue i , is sent to queue $j = R(x)$ for some function $R : \mathcal{S} \rightarrow [0, d]$ (if $j = 0$ the client leaves the system), and the client is either lost (CL) or restarts its service (RS) if the destination queue is full. Such event can



(a) Transition rate matrix



(b) Corresponding Markov automaton

Figure 14: Representation of the transition rate matrix (a) of the multiserver M/M/C/C queue, and its corresponding description as a Markov automaton (b) using C piecewise ASHEs a_1, \dots, a_C , each being drawn with rate μ .

be described as a piecewise ASHE over at most d zones $Z^k = R^{-1}(k)$ for $k = 1, \dots, d$ with the event on Z^k being an ASHE with direction vector $e_k - e_i$ and blocking relation $\mathcal{R}^{CL}\{(i, k)\}$ or $\mathcal{R}^{RS} = \{(i, k), (k, i)\}$ depending on the blocking policy. Provided the zones are polytopic, it is then easy to compute over-approximation of envelopes for this event. For some non-polytopic zones Z , the interval $\llbracket m, M \rrbracket \cap Z$ can still be easily computed with had-oc methods. For instance, if we consider the routing by index $R(x) = \arg \min_{i \in \{0, \dots, d\}} f_i(x_i)$ for some strictly increasing real functions f_i (we assume the f_i functions are chosen such that the arg min is unique for any $x \in \mathcal{S}$), then the corresponding zone is of the form $Z^k = R^{-1}(k) = \{x \mid \forall j \in \{1, \dots, d\}, f_k(x_k) \leq f_j(x_j)\}$, and one can show that $[m', M'] = \llbracket [m, M] \cap Z^k \rrbracket$ is given by $m'_k = m_k$, $M'_k = f_k^{-1}(\min_{i \in \{0, \dots, d\}} f_i(M_i))$, and $\forall i \neq k, m'_i = f_i^{-1}(f_k(m_k)) \vee m_i$ and $M'_i = M_i$.

Of course, routing policies can be combined with all the previous types of event (forks and joins, negative customers, batches, multiple servers) to give more types of piecewise ASHEs.

7.4 Application to a network

Here is an example of a network where events are ASHEs or piecewise ASHEs over polytopic zones. The following model permits us to compare the routing policies JSW to the Random (Rand) routing.

In this model, we consider two parallel queues 1, 2 with same capacity $C_1 = C_2 = 20$ and one server with service rate μ_1, μ_2 , respectively. We introduce $\mu = \mu_1 + \mu_2$ and $a = \mu_1/\mu$ such that $\mu_1 = a\mu$ and $\mu_2 = (1 - a)\mu$. Clients arrive in the system at a rate $\lambda = 1$ and go in one of the two queues, according to the chosen routing policies.

To compare the two policies, the 2-queue system is duplicated to give

a 4-queue-network, with two 2-queue subsystems with coupled arrivals and departures (see Figure 15). Incoming clients are duplicated via a fork and sent to each subsystem. If one of the subsystem rejects its copy of the client, the other must reject it as well. This is to ensure that both systems always receive the same quantity of clients.

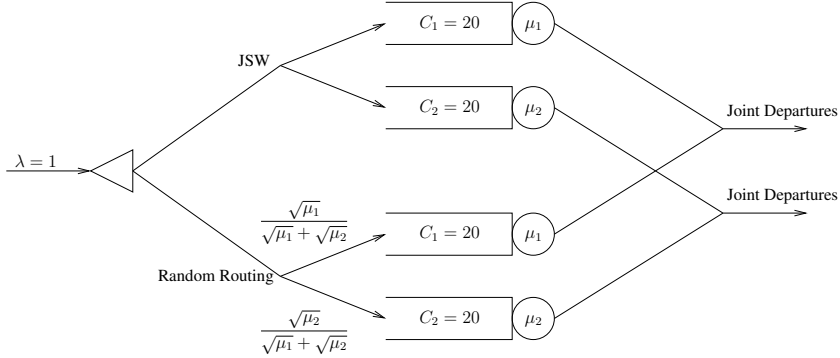


Figure 15: Network to compare the JSW and the Random routing policies

The first system then routes its clients using the JSW policy, while the second system sends its clients randomly to queue 1 with probability $\frac{\sqrt{\mu_1}}{\sqrt{\mu_1} + \sqrt{\mu_2}}$ and to queue 2 with remaining probability. If the packet arrives in a full queue, it is rejected (and the client is rejected by the other subsystem as well). The departures are also coupled: with rate μ_1 [resp. μ_2], a packet leaves queue 1 [resp. 2] (if the queue is not already empty), *simultaneously for each subsystem*. When the queue is empty in one subsystem, the corresponding queue in the other subsystem can still serve its clients.

Figure 16 shows the results of the simulation of that network using the envelope technique. We plotted the average difference of clients in the queue between the two subsystems, when the ratio a varies between 0.1 and 0.9, and for several values of the network load $\rho = \lambda/\mu$.

As we can see, the JSW routing strategy performs better than the random routing in all cases, and particularly when the serving rates are not close from each other or when the workload is high.

The coupling time is much lower for low load. In that case the dominant events are non-blocking services. If the load is high, the dominant events are arrivals that are blocking. Further, the coupling can only be achieved by hitting the lower boundary of the state space (by emptying the system), so the coupling time grows exponentially with the load.

8 Conclusion

Perfect sampling can be a very efficient method to study the stationary behavior of a Markov chain. This efficiency relies on three conditions. The

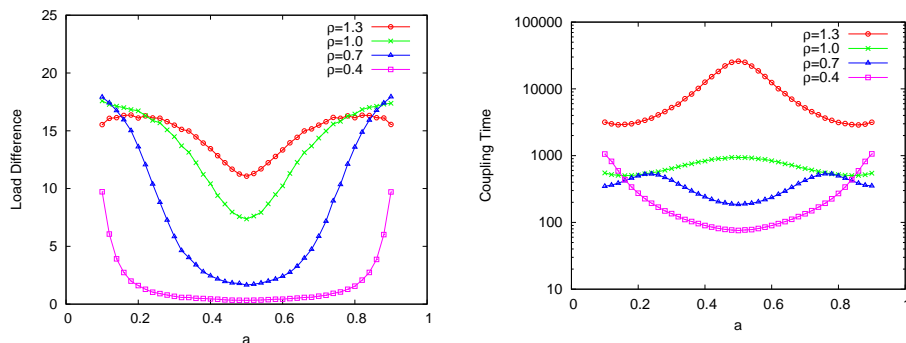


Figure 16: On the left: load difference between the two subsystems (Difference between the average number of clients in the random routing servers and the JSW servers) when a varies and for several value of the network load ρ . On the right: corresponding coupling time (number of iteration). Note that the scale is logarithmic in the number of iterations. For each point of the figure, 10000 simulations were run.

first one concerns the small number of simulated trajectories needed for unbiased coupling. The second condition is the requirement that a small time should be needed to compute those trajectories and the third condition is to have a small coupling time. When these three conditions are satisfied, the perfect sampling technique is hard to beat to compute the stationary distributed with a given confidence interval.

In this paper we show that the first condition is easy to meet by using envelopes. We also show that the second condition can be met for piecewise space homogeneous chains by using a special type of events (ASHEs) and by over approximating the envelopes. The third condition is more chain dependent. We provide several examples where coupling using the envelope technique happens fast and others where the coupling time is very large. In those cases, the splitting technique can help reduce the coupling time without compromising with the other two points.

Acknowledgments

The authors wish to thank Steven Oudot for his useful remarks.

References

- [1] F. Baccelli, B. Błaszczyszyn, and F. Tournois. Spatial averages of coverage characteristics in large CDMA networks. *Wirel. Netw.*, 8:569–586, 2002.

- [2] A. Bušić, B. Gaujal, and J.-M. Vincent. Perfect simulation and non-monotone markovian systems. In *Valuetools'08*, Athens, Grece, 2008.
- [3] C. G. Cassandras and S. Lafortune. *Introduction to discrete event systems*. Springer, 2nd edition, 2008.
- [4] J. Dopfer, B. Gaujal, and J.-M. Vincent. Bounds for the coupling time in queueing networks perfect simulation. In *Numerical Solutions for Markov Chains (NSMC'06)*, pages 117–136, Charleston, 2006. The 2006 A.A. Markov Anniversary Meeting (MAM 2006).
- [5] J.-M. Fourneau, I. Y. Kadi, N. Pekergin, J. Vienne, and J.-M. Vincent. Perfect simulation and monotone stochastic bounds. In *Valuetools'07*, Nantes, France, 2007.
- [6] B. Gaujal and F. Perronnin. Coupling from the past in hybrid models for file sharing peer to peer systems. In *Proceedings of the 10th International Conference HSCC'07*, Pisa, Italy, 2007.
- [7] B. Gaujal, F. Perronnin, and R. Bertin. Perfect simulation of a class of stochastic hybrid systems with an application to peer to peer systems. *Journal of Discrete Event Dynamic Systems*, 2007. Special Issue on Hybrid Systems.
- [8] O. Häggström and K. Nelander. Exact sampling from anti-monotone systems. *Statist. Neerlandica*, 52(3):360–380, 1998.
- [9] M. Huber. Perfect sampling using bounding chains. *Ann. Appl. Probab.*, 14(2):734–753, 2004.
- [10] N. Karmarkar. A new polynomial time algorithm for linear programming. *Combinatorica*, 4(4):373–395, 1984.
- [11] W. S. Kendall. Perfect simulation for the area-interaction point process. In L. Accardi and C. C. Heyde, editors, *Probability Towards 2000*, pages 218–234, New York, 1998. University of Warwick Department of Statistics, Springer-Verlag.
- [12] W. S. Kendall and J. Møller. Perfect simulation using dominating processes on ordered spaces, with application to locally stable point processes. *Advances in Applied Probability*, 32(3):844–865, 2000.
- [13] E. Oks and M. Sharir. Minkowski sums of monotone and general simple polygons. *Discrete and Computational Geometry*, 35(2):223–240, 1996.
- [14] J. G. Propp and D. B. Wilson. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms*, 9(1-2):223–252, 1996.

- [15] J.-M. Vincent. Perfect simulation of monotone systems for rare event probability estimation. In *Winter Simulation Conference*, Orlando, Dec. 2005.
- [16] D. B. Wilson. Perfectly random sampling with Markov chains. Annotated bibliography. <http://dimacs.rutgers.edu/~dbwilson/exact.html/>.
- [17] T. Zaslavsky. Facing up to arrangements: face-count formulas for partitions of space by hyperplanes. *Memoirs of the American Mathematical Society (Providence, R.I.: American Mathematical Society)*, (154), 1975.

A Proof of Theorem 4.4

Denote $[m', M'] = [m, M] \boxplus a$. Consider a fixed component j . We have the following different cases:

1. $j \notin \cup_{x \in [m, M]} B(x)$ (component j is not blocked in any state x , $m \leq x \leq M$). By Lemma 4.5, this is equivalent to $j \notin B(m) \cup B(M)$. Then:

$$[m'_j, M'_j] = [(m_j + v_j) \vee 0 \wedge C_j, (M_j + v_j) \vee 0 \wedge C_j].$$

2. $j \in \cap_{x \in [m, M]} B(x)$ (component j is blocked in all states x , $m \leq x \leq M$). It is easy to see that this occurs if and only if there is $i \in CR(m) \cap CR(M)$ such that $(i, j) \in \mathcal{R}$. Then $(x \cdot a)_j = x_j$, $m \leq x \leq M$ and:

$$[m'_j, M'_j] = [m_j, M_j].$$

3. $j \in (\cup_{x \in [m, M]} B(x)) \setminus (\cap_{x \in [m, M]} B(x))$ (component j is blocked in some but not all states x , $m \leq x \leq M$). Then we have the following cases:

- (a) There is $i \neq j$ such that $(i, j) \in \mathcal{R}$ and $i \in CR(m) \cup CR(M)$ (component j is blocked by a component i different than j).

If $v_j < 0$, then $(j, j) \notin \mathcal{R}$ or $j \notin CR(M)$ (remark that $(j, j) \in \mathcal{R}$ and $j \in CR(M)$ would imply $j \in CR(m) \cap CR(M)$ and thus $j \in \cap_{x \in [m, M]} B(x)$). We have:

$$[m'_j, M'_j] = [(m_j + v_j) \vee 0, M_j].$$

If $v_j > 0$, then similarly $(j, j) \notin \mathcal{R}$ or $j \notin CR(m)$ and:

$$[m'_j, M'_j] = [m_j, (M_j + v_j) \wedge C_j].$$

(b) Otherwise (only j is blocking j). Then $(j, j) \in \mathcal{R}$ and $j \in CR(m) \cup CR(M)$.

If $v_j < 0$, then $j \in CR(m) \setminus CR(M)$ (i.e. $m_j < -v_j \leq M_j$) and:

$$[m'_j, M'_j] = [0, (M_j + v_j) \vee (-v_j - 1)].$$

Note that $-v_j - 1 = \max\{x_j \mid x \in [m, M], j \in B(x)\}$.

If $v_j > 0$, then $j \in CR(M) \setminus CR(m)$ (i.e. $C_j - M_j < v_j \leq C_j - m_j$) and:

$$[m'_j, M'_j] = [(m_j + v_j) \wedge (C_j - v_j + 1), C_j].$$

For a given component j , the complexity of computing $[m'_j, M'_j]$ amounts to determining the corresponding case. However, this can be precomputed using $O(|\mathcal{R}|) = O(d^2)$ elementary operations. Thus the complexity of Algorithm 3 is $O(d^2)$.